

<https://helda.helsinki.fi>

---

## Understanding User Behavior in Naturalistic Information Search Tasks

Vuong, Tung

2019-11

---

Vuong , T , Saastamoinen , M , Jacucci , G & Ruotsalo , T 2019 , ' Understanding User Behavior in Naturalistic Information Search Tasks ' , Journal of the Association for Information Science and Technology , vol. 70 , no. 11 , pp. 1248-1261 . <https://doi.org/10.1002/asi.24201>

---

<http://hdl.handle.net/10138/307487>

<https://doi.org/10.1002/asi.24201>

---

cc\_by

publishedVersion

---

*Downloaded from Helda, University of Helsinki institutional repository.*

*This is an electronic reprint of the original article.*

*This reprint may differ from the original in pagination and typographic detail.*

*Please cite the original version.*

# Understanding User Behavior in Naturalistic Information Search Tasks

**Tung Vuong**

*Department of Computer Science, University of Helsinki, Pietari Kalmin katu 5, Helsinki, Finland.  
E-mail: vuong@cs.helsinki.fi*

**Miamaria Saastamoinen**

*Department of Computer Science, University of Helsinki, Pietari Kalmin katu 5, Helsinki, Finland.  
E-mail: miamaria.saastamoinen@gmail.com*

**Giulio Jacucci**

*Department of Computer Science, University of Helsinki, Pietari Kalmin katu 5, Helsinki, Finland.  
E-mail: giulio.jacucci@helsinki.fi*

**Tuukka Ruotsalo**

*Department of Computer Science, University of Helsinki, Pietari Kalmin katu 5, Helsinki, Finland and Helsinki Institute for Information Technology HIIT, Department of Computer Science, Aalto University, Konemiehentie 2, Espoo, Finland. E-mail: tuukka.ruotsalo@aalto.fi*

Understanding users' search behavior has largely relied on the information available from search engine logs, which provide limited information about the contextual factors affecting users' behavior. Consequently, questions such as how users' intentions, task goals, and substances of the users' tasks affect search behavior, as well as what triggers information needs, remain largely unanswered. We report an experiment in which naturalistic information search behavior was captured by analyzing 24/7 continuous recordings of information on participants' computer screens. Written task diaries describing the participants' tasks were collected and used as real-life task contexts for further categorization. All search tasks were extracted and classified under various task categories according to users' intentions, task goals, and substances of the tasks. We investigated the effect of different task categories on three behavioral factors: search efforts, content-triggers, and application context. Our results suggest four findings:

(i) Search activity is integrally associated with the users' creative processes. The content users have seen prior to searching more often triggers search, and is used as a query, within creative tasks. (ii) Searching within intellectual and creative tasks is more time-intensive, while search activity occurring as a part of daily routine tasks is associated with more frequent searching within a search task. (iii) Searching is more often induced from utility applications in tasks demanding a degree of intellectual effort. (iv) Users' leisure information-seeking activity is occurring inherently within social media services or comes from social communication platforms. The implications of our findings for information access and management systems are discussed.

## Introduction

Searching for electronic information has become a cornerstone of our everyday information-processing activities. People execute billions of web search queries every day,<sup>1</sup> and information searches are increasingly conducted across various applications and services beyond accessing the web.

Despite the importance of information searches, our understanding of users' search behavior has mostly been based on controlled laboratory studies (Jiang, He, & Allan,

---

Received May 26, 2018; revised November 28, 2018; accepted December 26, 2018

© 2019 The Authors. *Journal of the Association for Information Science and Technology* published by Wiley Periodicals, Inc. on behalf of ASIS&T. • Published online March 18, 2019 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/asi.24201

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

---

<sup>1</sup> On average, more than 6 billion daily searches are executed on web search engines  
(<http://www.internetlivestats.com/>)

2014; Mitsui, Shah, & Belkin, 2016; Orso, Ruotsalo, Leino, Gamberini, & Jacucci, 2017; Vakkari, 2003), interviews or diary studies (Church, Cherubini, & Oliver, 2014; Church & Smyth, 2009; Sohn, Li, Griswold, & Hollan, 2008; Teevan, Alvarado, Ackerman, & Karger, 2004), large-scale search engine log analysis (Broder, 2002; Jansen, Spink, & Saracevic, 2000; Silverstein, Marais, Henzinger, & Moricz, 1999), and log-analysis of naturalistic information-seeking behavior (Kelly, 2006a, 2006b).

Log studies rely on search engine logs that are quantitatively large and powerful for statistical analysis, but that lack contextual information beyond what can be observed within the search system. Meanwhile, diaries and interviews are limited to users' expressions and may suffer from interviewer or recall bias. Laboratory studies, on the other hand, allow for in-depth analysis, yet they often employ simulated work tasks that cannot reveal users' naturalistic search behavior. Kelly's (2006a, 2006b) research is closest to our study and reports naturalistic information-seeking behavior from data comprised of logs, interviews, and human assessments. However, it is limited to logs of online documents and concentrates on the effects of longitudinal variables, such as the endurance, frequency, and stage of seeking, rather than variables associated with the search behavior and immediate context of search tasks.

Therefore, although we have started to understand what users are searching, as well as where, at which stage, and how frequently they are searching within a broader task, our knowledge of the contextual factors affecting users' search behavior beyond what can be observed from document logs, self-reports, and interviews remain limited. These include the information that triggers users to search, the application and cross-system interactions that are associated with searching, and the search efforts that users are investing in longer search tasks. As it is well understood that information searching is dependent on the type of task that the user is completing (Byström & Hansen, 2005; Byström & Järvelin, 1995), we are interested in whether there are differences across various search task types. For example, these include differences in searches conducted to check facts, searches conducted as a part of a creative work, and searches performed in one's free-time to enjoy oneself.

To this end, our goal was to understand a user's naturalistic search behavior by investigating the effect of different real-life task categories for three behavioral factors: search efforts (how much and how long users search), content-triggers (how often the searches are dependent on content that users have already seen on their screens), and application context (what are the application types that form the cross-system interactions prior to searching). We studied the research questions with respect to various task factors: individual intentions (for example, being creative or checking facts), task goals (for example, communicating with someone or as a part of an intellectual work task), and substances (for example, free-time or programming).

In detail, we asked the following research questions:

- RQ1: Are there differences in content-triggering information searches with respect to different task categories: How often does the content that users have seen prior to searching trigger their searches?
- RQ2: Are there differences in application context in searching with respect to different task categories: What are the application types that the users are using prior to searching?
- RQ3: Are there differences in search efforts with respect to different task categories: How often and for how long do users search?

To answer the research questions, we report an experiment where naturalistic information search behavior is captured by analyzing a 24/7 continuous recording of information on 10 participants' computer screens captured over the course of 14 days. The participants' task diaries were used as the context in classifying search tasks under different categories and task factors. We then extracted various behavioral data describing content-triggers, application contexts, and search effort. We report significant differences across task categories in contextual and behavioral factors. Our findings revealed that:

- Search activity is integrally associated with the users' creative processes. The content users have seen prior to searching more often triggers a search, and is used as a query, within creative tasks.
- Searching within intellectual and creative tasks is more time-intensive, while search activity occurring as a part of daily routine tasks is associated with more frequent searching within a search task.
- Searching is more often induced from utility applications in tasks that demand a degree of intellectual effort.
- Users' leisure information-seeking activity is occurring inherently within social media services or comes from social communication platforms.

## Background

In this section, we discuss earlier research that focused on using the tasks as search context, task factors, and how they affect search behavior. We seek to inform and motivate our approach of studying search behaviors in naturalistic settings by analyzing continuous recordings of users' computer screens. In particular, we show that, although it is not possible to directly apply previous task categorizations to the study in naturalistic settings, the same search and behavioral factors coming from the literature including task goals, individual intentions, substances, triggers, search efforts, and application contexts are used in our study. Such an approach allows us to develop new task categorizations that match users' natural behavior and are compatible with previous work.

### *Tasks as Search Context*

Broader real-world tasks, as opposed to only search tasks, have been proposed to be the major contextual factor of information retrieval and seeking (Vakkari, 2003). Byström and Hansen (2005) conceptualized the tasks under two different views. A task can be an abstract construction

as defined in a task assignment, or can be viewed as concrete steps striving for a particular goal. The latter view is applied in the present study. A goal-oriented task may include smaller subtasks (Vakkari, 2003), and part of these subtasks can be information search tasks. Occasionally, the tasks have been replaced with search tasks that may, however, represent some underlying task context (Borlund, 2016). The underlying tasks and search tasks should be seen as nested rather than identical (Byström & Hansen, 2005). To make the tasks distinguishable from search tasks, researchers have begun to refer to them as work tasks (Byström & Hansen, 2005). The authors' intention was not to restrict the work task to some work context, but to include other activities within its scope. To avoid such confusion, we use the term *broader tasks* to cover all of the work tasks and leisure-related tasks.

Broader tasks and information retrieval (IR) are highly interconnected, as shown in many prior studies (Byström & Hansen, 2005). Ingwersen and Järvelin (2005) proposed a model of information access where task goals, task processes, information need and use, and information searching, as well as information systems, were correlated and were bound in a complex interaction. The findings of the empirical study by Pharo and Järvelin (2004) suggested that broader task goals directly influenced the information-seeking process, and the authors also stated that the characteristics of broader tasks should be considered when analyzing user search behavior. Similarly, Vakkari (2000) has shown that different process stages of a broader task were connected to the type of needed information, search terms, tactics and relevance criteria.

In research settings, broader tasks are either simulated or naturally occurring within a field study setting. Simulated tasks in IR experiments are artificial constructs assigned by the experimenters for the purpose of a specific research design (Byström & Hansen, 2005). The simulated tasks differ from the naturalistic real-life tasks in that they can be systematically varied and several variables can be controlled (Byström & Hansen, 2005; Xie & Joo, 2012). The fieldwork, however, has a few number of studies on broader tasks' effect on IR in naturalistic settings (Kellar, Watters, & Shepherd, 2007; Kelly, 2006a, 2006b; Kumpulainen, 2014; Kumpulainen & Järvelin, 2010). These studies are longitudinal and apply methods that capture data from both work and leisure time of the participants. The participants were asked to pair their tasks with the documents that were used, but the analysis concentrated on the within-user development of search behavior rather than on more generalizable task-wise differences (Kelly, 2006a, 2006b). Kellar et al. (2007) studied real-life search tasks, but their connections to broader work or leisure tasks were not recorded or analyzed. If naturalistic search behavior is studied, the data are often the participants' self-reports that include interviews (Teevan et al., 2004), diaries (Church & Smyth, 2009; Sohn et al., 2008), or questionnaires (Kumpulainen & Järvelin, 2010), which involved ex post facto rationalizations and other limitations on the data, such as server-side logs (Jansen et al., 2000; Silverstein et al.,

1999), web-based interaction logs (Kumpulainen & Järvelin, 2010), and limited operating system (OS) logs (Kelly, 2006a, 2006b). Obtaining search tasks' natural context that is their underlying broader task based on these logs is difficult. The present research belongs to the naturalistic field study and promotes a more naturalistic approach in data collection via continuous screen recording. Screen recording allows for the direct observation of some parts of user activity that can be hard to capture in conventional logging systems, such as the content that was appearing and applications that were used before, during, and after search.

### Task Factors

Tasks have been categorized according to various task features in the literature. Two main approaches have been proposed: theory-driven categorizations have been constructed based on combining and analyzing several empirical and theoretical studies (Li & Belkin, 2008), while data-driven categorizations have often been used in the studies that first put the focus on the participants in the specified context of information seeking and only then attempt to analyze the tasks based on a grounded theory methodology (Hansen, 2011). Li and Belkin (2008) made a thorough review of theory-driven task types. While these categorizations provide a variety of different viewpoints, they are often bound to a particular domain or a homogeneous group of users who participated in a particular study. For instance, Whitley and Frost (1973) focused only on research laboratory tasks, and Byström and Järvelin (1995) investigated tasks in municipal administration tasks. In these studies, the search tasks were carefully designed by experimenters to allow for the testing of certain hypotheses in controlled conditions. These analyses are by no means representative of general tasks and may not reflect naturalistic information search behavior.

Therefore, the homogeneity of the participants and their search context in these studies can be considered as an advantage, but that might also lead to a fallacy that studies about a homogeneous group and context could be generalized to other groups or contexts. As the present study is an exploratory study that analyzes domain-independent tasks, we had no initial reason to expect that the tasks should directly follow any predetermined theory-driven categorization. For categorizing tasks, previous works have been based on the following common factors: *task goals*, *individual intentions*, and *substance domain* of the tasks.

**Task goals.** Goal-driven task categorizations use the output target of the task as the factor (Rose & Levinson, 2004). The previous work has also proposed data-driven categorizations that do not include any domain-specific task types, and hence are broadly applicable to other domains as well (Saastamoinen & Järvelin, 2016). The categorization is task goal-driven and particularly suited for studying naturalistic search tasks. It aims to derive categories by seeking an answer to the question: "What goals are

the users trying to achieve in the task?” Examples of goals are whether the user is trying to communicate information or attempting to learn or achieve intellectual targets.

*Individual intentions.* While task goals determine the expected outcome of the search task, individual intentions behind the tasks influence the search process. People searching for information related to their hobbies or work can be driven by different individual intentions even though they would aim for a similar goal.

For instance, self-report methods refer to diaries that have been used in many field studies and the data are the representations of searching rather than searching process itself (Kumpulainen, 2014). Without a doubt, these methods give valuable information, especially about participant intentions and thoughts, as these cannot be directly observed. We followed the abstract concept of the everyday life information-seeking model (Savolainen, 1995). The individual intentions factor refers to preferences given to a task based on the individuals’ choices in everyday life, thereby answering the following question: “What individual intentions are the tasks serving?” The individual intention classification divides things into diverse groups according to their value to the searcher.

*Substance domain.* A third often-used source for categorization is the substance domain, which answers the following question: “What is the main domain that defines the task?” This factor has been particularly used in modeling information seeking for one specific professional group in one study; for example, nurses (Johannisson & Sundin, 2007), vault inspectors (Veinot, 2007), clergy (Dankasa, 2017), or researchers (Kumpulainen & Järvelin, 2010), and city administration (Saastamoinen, Kumpulainen, & Järvelin, 2012). For instance, all business-related tasks belong to the same substance domain of business regardless of their goal or intention. Task categories regarding the substance factor are mutually exclusive, which means that every task must belong to only one substance category. However, in actuality, a task may have the features of several substance categories. For example, a studying and researching task may be related to programming work. In these cases, the category was separated and we selected the category that was more emphasized by the participant’s task description. For example, in the case of programming tasks, all programming tasks were separated under a new category.

Task goals, individual intentions, and domain substances are all data-driven. The categorizations stemmed from the data and are independent from any a priori task structure. Naturally, the basis of task categorization may shift from data-driven to theory-driven, or vice versa, during the analyzing process. The researcher may begin to interpret the data to find meaningful classes and only then realize that a ready-made categorization suits perfectly; or the classes derived from theory prove to be useless and must be rearranged. As the categorizations are data-driven, the clusters are only named afterwards and the labels are as suitably selected as possible.

## *Context and Search Behavior in Information Seeking*

Research exploring context at the information-seeking level is concerned with the goals, intentions, and tasks that information seekers are trying to accomplish, but also the interaction with information to resolve the tasks and goals (Kelly, 2006a). Information-seeking behavior has been shown to be associated with the conventional behavioral measures, such as how much time the user spends seeking information, or the task stage in which the information seeking is conducted (Kelly, 2006a, 2006b; Li & Belkin, 2008). Less attention has been devoted to understanding the factors in the information or *application context* that influence when users need to search, how the informational contexts *trigger queries*, and how much *search effort* the users are investing in the search. Such evidence would enable bridging among the types of tasks and the type of informational resources that are being used in the search context (Kumpulainen & Järvelin, 2010). We set out to understand the following contexts and behavioral factors that are related to naturalistic search behavior.

*Content-triggers.* This factor refers to the contextual information that triggers the search process based on the evidence obtained from the connection between the selection of query terms and the content that users have seen prior to the search. Previous research has shed light on the context, search tactics, and search behavior, and their dependencies on task factors. For example, Vakkari (2000) analyzed the choice of search terms that were used in different stages of a task. Spink, Wolfram, Jansen, and Saracevic (2001) focused on the modification of queries within the search sessions and Eickhoff, Dungs, and Tran (2015) studied which concrete terms the user observed on the webpages and search engine result pages (SERP) that triggered a search and, subsequently, were reused as query terms. Eickhoff et al. (2015) found that a high share of newly added query terms were indeed previously seen by the user on SERPs. The authors interpreted this observation as evidence of query term acquisition, but the user study was based solely on the web and took place in a controlled lab environment.

*Application contexts.* A factor that is directly related to the content-triggering information is an application. Most of the applications are designed with a user interface to facilitate interactions between the users and the content displayed on the interface’s window, thus making the applications and the content-triggers closely coupled. Kumpulainen and Järvelin (2010) studied the applications used in information-intensive tasks that contained search tasks. They found that search tasks are interwoven with complex between-systems transitions of a variety of applications. These applications can be an important source of information needs, but also a source of inspiration for triggering searches and formulating queries. While the study provides initial evidence regarding the importance of the application context and between-systems transitions, it is limited to a molecular medicine domain.

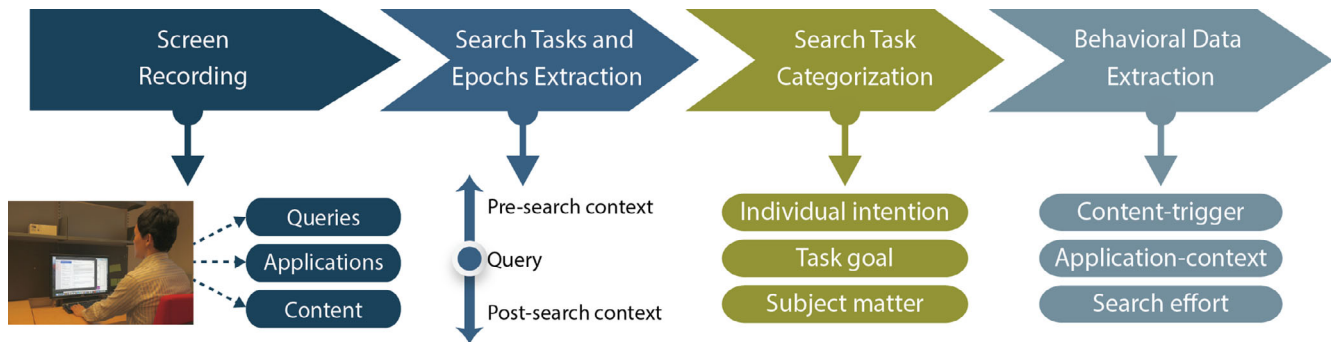


FIG. 1. The experiment procedure consisted of four phases: screen recording for a duration of 14 days, search tasks and associated search epochs extraction, search task categorization according to different task factors, and search behavioral data extraction. [Color figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

Consequently, the evidence on the contextual factors in the informational and application context that vary across naturalistic search tasks remains unknown. We focus on revealing the effects of contextual and behavioral factors. We study whether query terms can be linked to the content that the users observed on the screen prior to commencing a search. We refer to this as a content-trigger. We also study the transitions from different applications when queries are initiated by identifying the applications used prior to a search, and we refer to this information as the application context. The screen recording approach also allows us to identify the content and within-application transitions (information changes on an application's interface), which cannot be systematically revealed by using traditional logging.

*Search effort.* Finally, we study search effort, which is widely used as a dependent factor to measure task performance in IR studies (Gwizdka & Spence, 2007; Li & Belkin, 2008). While variables for measuring search effort have been previously studied using search engine logs (Cole, Gwizdka, Liu, & Belkin, 2012; Saastamoinen & Järvelin, 2016), we are able to identify queries within any application, including searches performed in desktop applications, custom search functionality embedded in web applications, and unconventional interactions that can be interpreted as a search, such as using pointing in geographic maps to acquire information.

## Methods

To capture naturalistic search behavior, we used a methodology in which the participants' computer screens were continuously recorded and all information appearing on their screens during the data recording period was captured. The screen recordings were then analyzed through a semiautomated process where we used automatic methods for extracting search tasks and epochs, which were manually corrected, verified, and categorized in the selected task categorizations. The behavioral data were then extracted to characterize each search epoch. The experiment procedure is presented in Figure 1 and is described in greater detail in the following sections.

### Screen Recording

Unlike other logging methods that have been used in traditional IR studies, screen recording is not restricted to prespecified logging functions. Apart from voice search, it can capture all possible search activities, including every input, as well as the presentation of content on the screen that occurs between the user and the computer before, during, and after searching. The screen recording approach was introduced in the previous work (Vuong, 2017).

*Apparatus.* We used a video screen recorder to record the screen frames of active windows at 5-second intervals or screen frames indicating information changes on the screen. In addition, we also collected OS log information that is associated with the screen frames, including the titles of active windows, the names of active applications, the Uniform Resource Locators (URLs) of webpages on active applications, and timestamps.

The video screen recorder was developed in two versions: We used Core Graphics API to implement the Mac OSX version, and we used Desktop App UI to implement the Microsoft Windows OS version. Both perform identical functions, recording and saving screen frames as images, along with the aforementioned OS log information. To produce a textual representation of the content on a screen frame's image, we used Tesseract (version 3.04), which was a very accurate optical character recognizer (OCR) engine.

*Participants.* Ten participants in both university and industrial settings with varying professions took part in the study. They were university students, computer scientists, entrepreneurs, and accountants. Five participants were males and five were females. There were seven people with master's degrees and three people with bachelor's degrees. The average age of the participants was 28 years ( $SD = 6$ ). The participants also had different cultural backgrounds: four were from Nordic countries, one was from central Europe, one was from Eastern Europe, and four were from Asia.

The participants were recruited via a posting that was distributed to mailing lists. A questionnaire was attached to the recruiting message, which was sent to the relevant mailing

lists to collect background information on potential candidates. Only respondents who used a laptop as their main device for performing their everyday digital activities were considered to be eligible for the study. Another eligibility criterion for taking part in the study was having a higher education background, as we assumed that people satisfying this criterion would more likely use their laptops for everyday digital activities.

The participants were informed of our privacy guidelines prior to joining the experiment and were told that the video recordings were stored on their own computers during the screen recording phase. After that, the data would be transferred to a secured server and used only for research purposes. After the experiment, the participants were compensated with three movie tickets worth ~40 USD. A consent form was obtained from the participants regarding the data usage, privacy, and procedure of the experiment. Although the data collected from the same group of participants were reused in the present study, only a subset of the data was used and the study was distinct from the previous work (Vuong, 2017).

**Procedure.** Upon agreeing to take part in the experiment, the video screen recorder was installed on each participant's laptop and set to run continuously in stealth mode for a duration of 14 days. A screen recorder was automatically launched whenever the laptop was turned on. During the recording phase, the participants were advised to use their laptops as usual and to avoid stopping the recorder unless it was necessary.

After the installation of the video screen recorder, the participants were each asked to keep a diary of their daily tasks. For the convenience of writing a diary, we provided the participants with a diary template including three fields: a brief statement describing the task, specific keywords related to the task, and the names of the available people involved in the task. The participants used pen and paper to write in the diaries, and they could write the diaries whenever they felt comfortable throughout the day. We intentionally advised the participants to focus on writing a broader task consisting of several activities. We encouraged the participants to use their own conceptual understanding of what activities could make for a meaningful broader task.

After the 14-day period, the participants visited our lab and the video screen recorder was uninstalled from their laptops. We skimmed through every written task, marked the tasks that had unclear descriptions, and discussed them in detail with the participants. During the discussion, the participants were allowed to refine their diaries by adding any missing tasks that they wanted, combining task entries regarding the tasks they felt were too specific, or separating tasks that were too broad.

### *Search Tasks and Epochs Extraction*

A search task includes a query or several queries and has a uniform motivation, or an information need that evolves seamlessly in the work flow of a diary task as a motivation for conducting immediate search activities. To effectively identify naturalistic search tasks, we decomposed a search

task into one or several search epochs. Each search epoch contained a query that a user submitted to the search engine and the associated presearch and postsearch context. To determine that multiple continuous search epochs belonged to the same search task, we used the corresponding task in the diary as the context for understanding whether several search epochs shared the same search goal and belonged to the same broader diary task.

A wide spectrum of search tasks was extracted in the experiment. For instance, we extracted local file search activities using the OS-specific applications, such as Finder, Spotlight, and Explorer. We also recorded searches using map interfaces, such as Google Maps, with typed queries, drags, clicks, and searches in email clients, as well as custom searches on websites.

**A search epoch.** The preliminary step of our analysis was to detect search epochs from the screen recordings. Figure 2 (part 3) illustrates a search epoch from a participant's screen recordings. A search epoch comprised three parts: a query frame, presearch context, and postsearch context.

- *A query frame* is a reference frame of a search epoch. It is a screen frame of an SERP that was recorded in response to a query issued by a searcher. The regular expressions in the Appendix<sup>2</sup> were applied to find all candidate query frames in the participants' screen recordings.
- *Presearch context* is a temporal sequence of screen frames recorded at 2-minute intervals prior to the query frame. In the case of missing screen frames due to the computer being idle within the presearch context, we extracted one frame that temporally precedes the query frame.
- *Postsearch context* is a temporal sequence of screen frames recorded at 2-minute intervals subsequent to the query frame. Similarly, when there are no existing screen frames in the postsearch context, one temporally successive screen frame from the query frame was extracted.

**A search task.** Based on the determined search epochs from screen recordings in the previous step, we formed a search task consisting of a set of search epochs. Figure 2 (part 2) illustrates how a search task is formed. Search epochs can, but do not have to, follow each other temporally. In other words, a search task can be one isolated search epoch when the presearch context and postsearch context do not overlap with subsequent search epochs. In another case, several continuous search epochs sharing the same information need are combined as a search task. Two experts manually corrected the extracted search tasks by removing collections of screen frames that were not search tasks, then separating a sequence of search epochs into several search tasks if they belonged to distinct search tasks.

### *Search Task Categorization*

Unlike search tasks in laboratory settings, a naturalistic search task is not a self-standing search assignment; rather,

<sup>2</sup> <https://tinyurl.com/ybcyasd4>



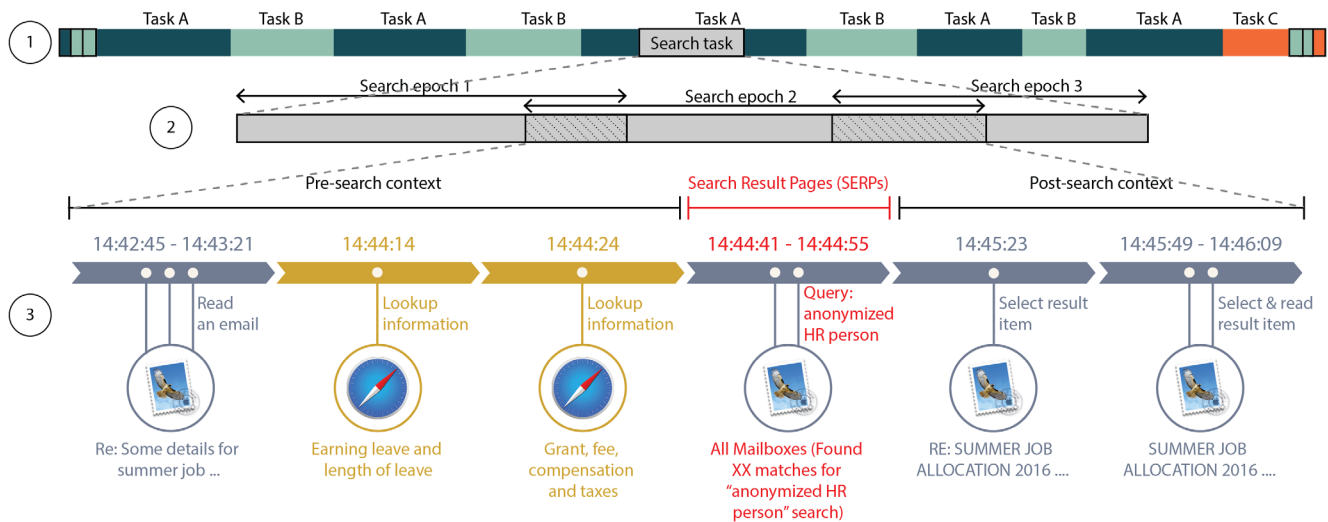


FIG. 2. The granularity of a diary task that contains naturalistic search tasks and associated search epochs. (1) A user's digital activities include several diary tasks. (2) A search task is composed of three search epochs. (3) An overview of a search epoch. In the presearch context, a user engaged in a search task involving verification of factual information regarding human resource policies to reply to an email from a new employee. A query frame is the screen frame containing the user's issued query. In the postsearch context, a user used the retrieved information to respond to the email. [Color figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

it is strictly dependent on the broader task at hand. Thus, we used the diary task as the broader task context to correctly categorize naturalistic search tasks. Manual categorization for every search epoch for a search task and every search task for a corresponding task category can be error-prone and laborious because the data are large. To overcome this, we designed a two-phase semiautomatic procedure for categorizing search tasks.

In the first phase, we used latent semantic analysis (LSA) (Deerwester, Dumais, Furnas, Landauer, & Harshman, 1990) to uncover the topical structure of the OCR-processed screen frames. LSA learns a latent lower-dimensional representation of the input data. Each dimension in the lower-dimension space can be interpreted as a topic, and each topic can be associated with a diary task. Each topic assignment was manually verified and, if necessary, corrected by two expert annotators.

In the second phase, each search task and the associated search epochs were mapped to the diary task identifiers that were defined in the previous phases. To ensure the reliability of the annotations, a double-blind annotation was performed by two expert annotators. Cohen's kappa test showed high agreement between the annotators ( $\kappa = 0.85$ ). Then the expert annotator formed the categorizations to all diary tasks in the data. The task category classification was based on the written descriptions of the tasks given by the participants. The task categorizations were data-driven; thus, the expert annotator considered features that were shared between the diary tasks and clustered similar diary tasks according to the three task factors: individual intentions, task goals, and substances of the tasks. The task categorizations were rather abstract and, hence, were usable across the tasks conducted by a range of knowledge workers in the study. Despite the inherent

problems with the loss of specificity of the categories, it was necessary to be able to create categories that were general enough to represent the data and to allow for conclusions to be drawn based on the categorization.

After forming stable clusters, the expert characterized their features and labeled them with descriptive titles. It is important to note that those features are clearly stated in the task descriptions written by the participants, and were relatively straightforward to categorize. Task factors, the corresponding categories, and examples are shown in Table 1 and some examples are discussed in the following paragraphs.

Four categories were formed under the Individual Intentions factor: (i) Tasks with the intention of Being Creative shared the two dominant features, which were writing/composing documents. (ii) Tasks with the intention to Enjoy Oneself shared two common features, which included social media activity and video streaming/music listening. (iii) In Gain Knowledge, the tasks were described with the two features of learning and research-related activity. (iv) The rest of the tasks fall into the Daily Activity category. These tasks represent a variety of routine activities, such as continuously making travel plans/accommodation arrangements, online shopping / daily e-commerce, following up-to-date news, and managing personal information.

The Task Goal categorization adopts the earlier categorization (Saastamoinen & Järvelin, 2016) and is based on the following generic features. (i) Tasks with a Communication goal have the main feature of communicating with other people as the precondition for success within the task. These can include going through email conversations and replying to the messages, or taking part in a live video call. (ii) The Maintaining/Advancing category has the feature of whether the task is at the core of the substance of the work or, rather, that supports a main function. These are typically information



TABLE 1. Task factors, the corresponding task categories, and examples are taken from the diaries.

Task factor	Task category	Examples
Individual intentions	Be creative	Essay writing-marketing, promotion...; Proactive search simulation analysis.
	Enjoy oneself	Watching movies and TV shows; Browsing an online forum for knitting and crochet.
	Gain knowledge	Reviewing discriminative representation learning; Foreign language studying with two teachers.
	Daily activities	Trying to find accommodation in a city; Following up latest news on BBC, CNN.
Task goals	Communication	Writing emails to a potential summer trainee; Arranging a job interview at a company.
	Maintaining and advancing	Updating a to-do list for a project; Starting ICT company papers.
	Seeking or receiving information	Finding scientific venues; Finding and listening to music.
	Intellectual	Studying C++; Preparing a pitch deck for a startup.
Substances	Free-time	Watching "Lost" seasons; Listening to music on a QuickTime player.
	Business	Generating business calculations; Writing an article about a startup with the entrepreneur.
	Programming	Modifying a user interface; Extracting keywords from software with KEA.
	Social life	Looking for friends on Facebook; Checking Slack.
	Studying and researching	Reviewing MTAP paper Face super resolution based on...; Reading LSTM RNN recurrent neural network.

searches for administrative tasks or tasks where an expected larger output is approached gradually. They were easily recognizable from the task descriptions with "reviewing," "starting something new," or "continuously updating a document." (iii) Seeking or receiving information are tasks that aim to acquire a specific piece of information by actively seeking it, or passively receiving it. The diary entries corresponding to this goal often began with "finding something," "looking something up," or "watching something." (iv) Tasks with the Intellectual goal has the feature of demanding a degree of intellectual effort.

Categories in the Substance task factor reflects the domain substance of tasks in the data. These include five general categories: (i) Free-time; (ii) Business or industrial job-related tasks: these excluded tasks in the academy; (iii) Programming tasks' scope can be the whole process of software development, not just coding or scripting; (iv) Social life tasks mostly involved social media activity; and (v) Studying and researching tasks can be academic or industrial research and development tasks.

After the categorization, a double-blind interrater agreement was determined by asking another researcher, who had no prior knowledge about the study or the task categories, to do the categorization. The interrater agreements were found to be high for all categories (Cohens kappa, Individual Intention: 0.72), (Cohens kappa, Task Goal: 0.81), (Cohens kappa, Substance: 0.88).

### Search Behavior and Contextual Factors

We analyzed three factors related to naturalistic search behavior: (i) content-triggers: to understand whether the content the users have seen prior to searching that triggers the search; (ii) application context: to understand which application category that the users have frequently used prior to searching; and (iii) search efforts: to determine how much effort the users spent on search tasks and which task categories involve more searching than others.

*Content-triggers.* Content-trigger refers to a presearch context that contains any of the query terms. Content-

triggers were extracted by comparing the keywords of the query with the content the users had seen on the screen frames in the presearch context. A program was implemented to automatically extract whether any term in the set of keywords that existed in the content of the screen frames in the presearch context. Taking a search task in Figure 2 as an example, the query "anonymized HR person" was submitted to Mail's search interface. The phrase of this query originally appeared on the screen in a presearch context that triggered user search activity.

Content-triggers that were a combination of stopwords were discarded. During the process, we also noted that several keywords featured a set of stopwords, but were meaningful in the presearch context. We further manually checked and verified the correctness of individual content-trigger.

*Application context.* The application context refers to an application used prior to searching. We used OS log information associated with the screen frames to extract the name of the application that was used before a query frame. The applications were locally installed applications on the participants' laptops. Due to the large number of occurrences with respect to various web browsers, we decided to extract the domain names of the webpages visited by the participants and considered them to be separate application context.

The applications used by the participants were very diverse. We further manually categorized applications into seven groups based on their common functions, types, and fields of use. The application categories are presented in Table 2. The Social category included applications where the main function was to enable communication with other people (for example, Skype, Mail). General web search engines and information sources (for example, Google, Wikipedia) were categorized into the Search Engine category. Participants used many dedicated tools to support their search tasks (for example, various digital dictionaries), and these were categorized into the Support category. The Transactional category typically featured websites used to support interaction and even to enable transactions (for example, online stores, journey planner). Meanwhile, the Static category included static websites that did not support interaction or transactions

TABLE 2. Application categories are data-driven. They are categorized based on common function and type of use.

Application category	Description
Social	Applications and websites where the main function is to enable communication between people.
Search Engine	General web search engines form a category of their own. Users re-visited the search engine application after some time to perform new searches.
Support	Applications or language support tools that support the search task.
Transactional	Websites that are typically used for manifold interactions and that support interaction and even enable transactions.
Static	Static websites that are typically used for browsing and that do not support or encourage much other interaction or transactions.
Local	Local applications that are installed on the participant's computer.
Other	Rare websites are placed in this category.

(for example, personal weblogs or online tutorials). Locally installed applications were also grouped as the Local category, but this category excluded applications that were categorized into the aforementioned categories. For instance, the main function of instant messaging applications was to socially interact with others; thus, we classified it as the social category. Finally, any website that rarely occurred in the recorded data was placed into the Other category.

*Search efforts.* To understand some task categories that require more search effort than others, we extracted the number of queries submitted to a search engine in a search task, the duration spent on a search task, and the number of search tasks performed in a diary task.

For every search task, the number of queries was computed during search task extraction. An automatic procedure was used to count the number of queries for each search task based on the manually annotated data. The duration spent on a search task was computed in seconds by comparing the difference between the first screen frame and the last screen frame of a search task. Finally, for every reported task in the diary, the number of associated search tasks was automatically computed.

### Measures

The following set of measures was defined to operationalize user behavioral factors.

*Content-trigger.* A binarized variable was used to characterize whether a search epoch contained a content-trigger. If the query appeared in the presearch context of the search epoch, we marked the search epoch as content triggering. We computed the frequency of content-trigger occurrences as a triggering ratio across task categories.

*Application context.* The application context was measured as the share of application category appearances

directly prior to searching across task categories. In the event that no applications were used within 2 minutes prior to searching, the application context was assigned to the application" search engine."

*Search efforts.* We quantified the search effort by computing three measures: the number of queries per search task, the duration of search task, and the number of search tasks per diary task. For each measure, we report three corresponding values with respect to every task category: minimum, maximum, and average.

### Statistical Analysis

We applied the Kruskal–Wallis test to determine whether there were statistically significant differences in the behavioral factors among the task categories for individual intentions, task goals, and substances of the tasks. In the test of the significance levels of the task category differences, we used task category as the independent variable, and three behavioral factors as the dependent variables. SPSS v. 24 (IBM, Armonk, NY) was used for the calculation of statistical significance for three behavioral factors across task categories. All post-hoc tests were conducted with the Dunn test with Bonferroni correction.

## Results

The descriptive statistics of the recorded data are shown in Table 3. A total of 688 naturalistic search tasks were identified in the screen recordings. Overall, 49,647 screen frames were recorded for a duration of 14 days. Participants reported 119 diary tasks, and 69 diary tasks containing search epochs that were analyzed in the study. There were 1,299 search epochs belonging to the search tasks that consisted of 8,806 screen frames. We observed that 18% of screen frames belonged to the participants' search tasks during the 14-day screen recording period. Despite the large number of search epochs extracted from the recorded data, 42% of diary tasks (50 diary tasks) did not include any search tasks.

In the individual intention factor, the "be creative" category consisted of 132 search tasks, "enjoy oneself" contained 262 search tasks, "gain knowledge" included 182 search tasks, and "daily activities" consisted of 112 search tasks. For the task goal factor, there were 118 search tasks falling

TABLE 3. Descriptive statistics of the collected digital activity data from 10 participants.

	Total	Average
Number of diary tasks	119	12 ( <i>SD</i> = 2)
Number of screen frames	49,647	4,965 ( <i>SD</i> = 3,299)
Number of diary tasks containing search epochs	69	7 ( <i>SD</i> = 2)
Number of search tasks	688	69 ( <i>SD</i> = 45)
Number of search epochs	1,299	130 ( <i>SD</i> = 97)
Number of screen frames in all search tasks	8,806	881 ( <i>SD</i> = 720)

TABLE 4. The results of content-trigger were measured as a triggering ratio with respect to task categories.

Task factor	Task category	Content-trigger	P-value
Individual intention	Be creative	0.69	<b>.010</b>
	Enjoy oneself	0.60	
	Gain knowledge	0.55	
	Daily activities	0.60	
Task goal	Communication	0.56	.206
	Maintaining and advancing	0.67	
	Seek or receive information	0.59	
	Intellectual	0.61	
Substance	Free-time	0.63	<b>.001</b>
	Business	0.58	
	Programming	0.82	
	Social life	0.59	
	Studying and researching	0.57	

into “communication,” 86 search tasks falling into “maintaining and advancing,” 214 search tasks related to “seek or receive information,” and 270 search tasks for “intellectual.” Lastly, with respect to the substance of the task, 141 search tasks were in “free-time,” 68 search tasks were in “business,” 88 search tasks were related to “programming,” 188 search tasks belonged to “social life,” and 203 search tasks were included in “studying and researching.”

### Content-Triggers

Table 4 presents the frequency of content-triggers in different task categories. A statistically significant difference was found in the content-triggers for the categories of

individual intentions ( $\chi^2(3) = 11.42, p = .01$ ) and substances ( $\chi^2(4) = 17.83, p = .001$ ), but no differences were found in the case of content-triggers for the task goal categories ( $\chi^2(3) = 4.57, p = .206$ ). Follow-up, pairwise comparison tests showed a significant difference between “be creative” tasks and “gain knowledge” tasks ( $p = .005$ ), and between the “programming” task and the other four subject-matter categories ( $p < .02$ ).

A statistically significant difference was found in how often the search was triggered by content observed on the screen when it came to individual intentions. While users were “being creative,” information need was mostly triggered by the content showing a triggering ratio of 0.69. “Enjoying oneself” and “Daily activities” were on par with frequency showing a triggering ratio of 0.60. “Gain Knowledge” is less content-triggering, with a triggering ratio of 0.55.

Task goal was not dependent on the content-triggers, but substance was dependent. The “programming” category showed a significantly high frequency of content-triggers with a triggering ratio of 0.82. “Free-time” tasks had a triggering ratio of 0.63. “Studying and researching,” “business,” and “social life” tasks had fewer content-triggers with a triggering ratio of 0.58, 0.59, and 0.57, respectively.

### Application Context

Figure 3 shows the general results of the percentages of the seven application categories. Overall, applications in social categories are by far the most common applications that are used in the presearch context. Individual intentions

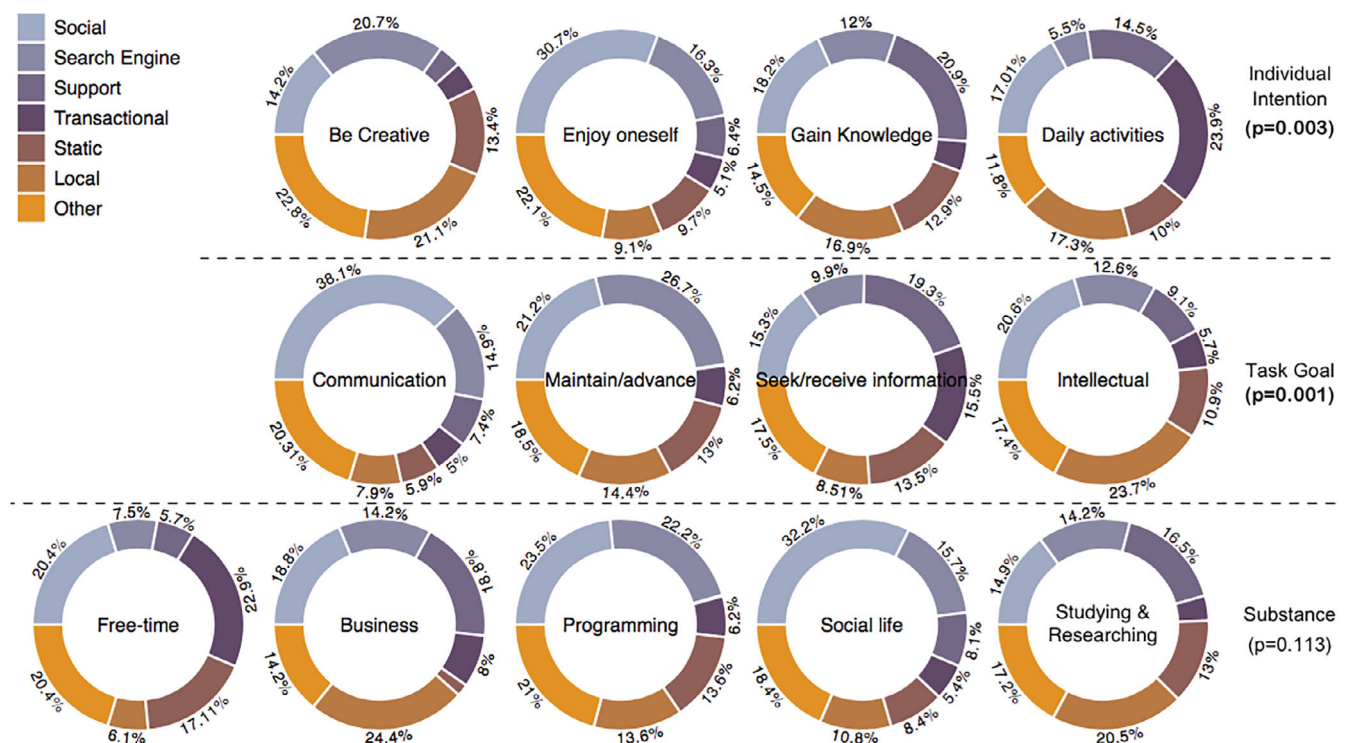


FIG. 3. Results of application context with respect to task categories. The number of occurrences of application categories prior to searches are measured as percentages. [Color figure can be viewed at wileyonlinelibrary.com]

TABLE 5. Results for search efforts measured as the number of queries per search task, duration of search tasks in seconds, and number of search tasks per diary task.

Task factor	Task category	Min-max	Number of queries per search task		Min-max	Search task duration		Min-max	Number of search tasks per diary task	
			Average	P-value		Average	P-value		Average	P-value
Individual intention	Be creative	1–6	1.8	<b>0.025</b>	103–2322	360	0.446	0–28	5.7	0.413
	Enjoy yourself	1–13	1.7		54–3515	448		0–58	10.9	
	Gain knowledge	1–10	1.8		62–3884	339		0–36	8.1	
Task goal	Daily activities	1–15	2.6		87–3987	416		0–31	6.6	
	Communication	1–13	1.7	0.139	96–2024	327	<b>0.026</b>	0–43	7.4	0.367
	Maintaining and advancing	1–9	1.7		92–1150	302		0–23	4.2	
	Seek or receive information	1–15	2.1		54–3884	490		0–58	10.7	
	Intellectual	1–9	1.9		62–3987	383		0–29	9.0	
Substance	Free-time	1–15	2.0	0.514	82–2368	365	0.697	0–34	6.9	0.475
	Business	1–9	2.6		121–3987	436		0–29	4.9	
	Programming	1–5	1.7		62–1542	346		0–23	4.8	
	Social life	1–13	1.8		54–3515	478		0–58	12.5	
	Studying and researching	1–10	1.8		67–3884	353		0–36	9.7	

and task goals were related to application context, with ( $\chi^2$  (3) = 13.75,  $p$  = .003) and ( $\chi^2$  (3) = 22.52,  $p$  < .001), respectively, whereas substances of the tasks were not significantly related, with ( $\chi^2$  (4) = 7.47,  $p$  = .113). Pairwise comparison revealed a significant difference between “enjoying oneself” and other individual intention categories ( $p$  < .013), and between “intellectual” and other task goal categories ( $p$  < .009).

For the individual intention factor, the “be creative” category had a highest percentage of using “other” applications as a prior application context, with 22.8%. Typically, while being creative, users mostly moved from rarely or single-time used applications to the search engine. For the “enjoy oneself” category, 30.7% of application context falls into the social application category. Users in the “gain knowledge” tasks mostly used support applications prior to a search, with 20.9%. Finally, transactional applications were mostly the application context prior to search in “daily activities,” with 23.9%.

For task goals, interestingly, “communication” tasks had the highest percentage of the application context from social applications, with 38.1%. For “maintaining and advancing” tasks, often revisited the SERP on search engines, with 26.7% and carried out a new search. The “seeking and receiving information” category had the highest percentage of moving from support applications to search engine, with 19.3%. “Intellectual” tasks mostly began with local applications (for example, a PDF reader) and moved to a search, with 23.7%.

### Search Efforts

Table 5 presents the results for search efforts for the selected measures. Overall, search efforts were related to the search task category, but the results suggest more fine-grained dependencies across the different task factors.

For individual intentions, search efforts measured as the number of queries per search task were dependent, with ( $\chi^2$  (3) = 9.37,  $p$  = .025), whereas the search task duration and number of search tasks were not dependent, with  $p$ -values of .446 and .413, respectively. For a search task in this categorization, an average of at least 1.7 queries were issued, and users spent an average duration of a minimum of 339 seconds (5.6 minutes) on the tasks, and they performed an average of more than five search tasks.

Search task duration was associated with various types of goals that affect how much time users spend on searching. Task goal was dependent on search duration, with ( $\chi^2$  (3) = 9.22,  $p$  = .026), but it was independent of the number of queries and the number of search tasks, with  $p$ -values of .139 and .367, respectively. Similar to individual intentions, on average a minimum of 1.7 queries were issued. The lowest search task duration, and the fewest number of search tasks in this categorization were reported with a minimum duration of 302 seconds (5 minutes); on average, more than four search tasks were performed.

Pairwise comparisons indicated a significant difference between “daily activities” and other individual intention categories in terms of search effort ( $p$  < .05), and between “intellectual” and “maintaining and advancing” in terms of the search task duration ( $p$  = .05).

Surprisingly, substances of the tasks had no effect on the search effort. This suggests that the substances of the task were not a determinative factor for search effort.

### Findings

The results also revealed interesting dependencies between the measured behavioral variables and the task categories. In the following, we distill generalizable findings

from the results and reflect on the research questions defined earlier.

**RQ1:** Are there differences in content triggering information searches with respect to different task categories? We found that content-triggering was associated with the individual intention and substance factors, but was not related to the task goal factor.

More precisely, when users searching for the task for “being creative,” their search was significantly more often triggered by the content that had appeared on their screens prior to commencing the search. Similar findings have been reported in the study of Eickhoff et al. (2015) using a lab-based analysis. However, in our study content triggering was mostly manifested while users were engaged in creative tasks. This suggests that, in naturalistic settings, the importance of content users have seen prior to searching is associated with a creative activity, such as writing or producing artifacts in some other ways, such as programming. Conversely, search tasks to “gain knowledge” that often involved reading and browsing document content seemed to be intrinsically evoked and were less reliant on content triggering.

**RQ2:** Are there differences in application context in searching with respect to different task categories? Task categories were found to be associated with different application contexts that appeared prior to searching.

Pairwise comparisons revealed that searching occurring in a social application context were induced by social applications more often when the task intent was to “enjoy oneself.” This finding is not surprising; it confirms the previous findings of Teevan, Ramage, and Morris (2011) that queries are often issued within social media services with leisure intent, such as finding celebrities appearing in streaming movies or music videos, locating people with similar interests, and navigating to friends’ pages to investigate social media activity. This suggests that users’ leisure information-seeking activity is occurring inherently within social media services or comes from social communication platforms.

Pairwise comparisons also revealed that search behavior in tasks with an “intellectual” goal, such as analyzing, researching, reviewing, and writing were more often induced from utility applications, such as word processing applications, spreadsheet applications, or programming platforms. Although not surprising, this suggests that intellectual tasks are strongly associated with applications that support knowledge work. This finding is in line with our findings with respect to RQ1, as a large portion of tasks with the intent of “being creative” were found to have an “intellectual” goal. Consequently, search behavior in creative and intellectual tasks was induced by the artifacts that the users were producing and occurred in the context of utility applications. This suggests that search activity is integrally associated with the users’ creative processes.

**RQ3:** Are there differences in search efforts with respect to different task categories? The number of queries and the investment of time across task categories were found to have significant differences between task categories.

Pairwise comparisons revealed that “daily activity” tasks were associated with larger number of queries than the tasks with intent to “be creative” and “gain knowledge.” This is in contrast to prior work suggesting that tasks including creative aspects were more search-intensive than daily routine tasks (Saastamoinen & Järvelin, 2016). A possible explanation is that past research was solely concentrated on the user work context and did not have access to users’ “daily activity” data outside the workplace.

Pairwise comparisons also revealed that tasks with an “intellectual” goal often took longer to complete, and the corresponding search tasks had a longer duration in comparison to the tasks with “maintaining and advancing” goals. Similar results have been recently reported by Saastamoinen and Järvelin (2018), thus showing an association with intellectual tasks and increased search activity. Our results suggest that such a correlation between the search task duration and the type of tasks, indeed, also occur in more naturalistic settings. Consequently, our results suggest that routine search activity is more search-intensive, but searching within intellectual and creative tasks is more time-intensive.

## Discussion and Conclusion

In this study, users’ naturalistic search behavior was studied using real-world data. Unlike any previous experiment that we are aware of, the data were collected using a naturalistic approach by simply recording the computer screen. The resulting data provided insight into naturalistic information behavior covering both leisure and work activities, and provided a view of all applications, including email, messaging, and utility software, as well as web activity beyond web search engines (for example, social media sites, e-commerce sites, and a variety of long-tail services).

### Implications

The implications of our findings are striking, as they reveal an interdependencies between broader contextual factors of digital environments and users’ information search and seeking behavior. Our findings demonstrate that there are differences in requirements and contexts across different types of tasks, substances of the task, and users’ internal intentions when completing information intensive tasks. We foresee implications for both researchers designing user studies and experiments, as well as for practitioners designing information access systems.

*Designing user studies and experiments.* Different task categories have significant impacts on the cognitive efforts in formulating queries and on how much time and effort users are investing in searching information. The content

triggering and the applications forming the context of users' digital activity before and after information search can inform us about specific types of simulated tasks that should be further studied and used in more controlled experiments. Given the knowledge of the importance of behavioral factors and task factors, our results can help researchers to design user studies and tasks for experiments that are more congruent with real-life behavior and can have better face validity "in the wild."

*Designing information access systems.* The effects demonstrated in our experiments can also inform designers and developers focusing to support particular types of tasks. Our results suggest that the current "one-size-fits-all" user interaction with search engines may not be optimal for different tasks and the design of the next generation of information access systems could benefit from considering broader user contexts to identify and even predict specific kinds of search support that might benefit the user. Our findings show that contextual factors and search effort that are linked to task categories constitute to a useful step in adapting IR environments. This can also be promising for customizing search by accounting for the importance of content triggering and applications used prior to search.

### Limitations

The main limitation of the present study was the limited number of participants. A smaller sample size, on the other hand, enables us to focus more on understanding individual participant's search tasks, associated intentions and goals, and particularly makes it possible to reliably determine application contexts and content-triggers affecting the search behavior in different types of tasks. Although the overall number of search tasks collected from screen recordings was relatively small, for example, compared to large-scale log studies, the results still revealed dependencies among the contexts and behavioral factors of different task types. Despite these findings, the reported study is exploratory and we have no intention of making claims about the generalizability of our finding to a larger population.

The recruitment setting included only young individuals with an average age of 28 ( $SD = 6$ ). Older or younger participants may have had different behavioral patterns when using digital tools and services. In addition, participants were all knowledge workers, which might also have affected the types of applications used and the task categories reported in the results.

Despite these limitations, the reported study on naturalistic search tasks is novel. There is no prior hypothesis formulation about user naturalistic search behavior, or suitable categorization schemes that could fit such recording data. The advantage is that we have 24/7 recordings of computer screens, which opens up new opportunities for insightful investigation about user behavior in naturalistic real-life information search tasks.

### References

- Borlund, P. (2016). Framing of different types of information needs within simulated work task situations. *Journal of Information Science*, 42(3), 313–323.
- Broder, A. (2002). A taxonomy of web search. *ACM SIGIR Forum*, 36(2), 3–10.
- Byström, K., & Hansen, P. (2005). Conceptual framework for tasks in information studies. *Journal of the American Society for Information Science and Technology*, 56(10), 1050–1061.
- Byström, K., & Järvelin, K. (1995). Task complexity affects information seeking and use. *Information Processing & Management*, 31(2), 191–213.
- Church, K., Cherubini, M., & Oliver, N. (2014). A large-scale study of daily information needs captured in situ. *ACM Transactions on Computer-Human Interaction*, 21(2), 10:1–10:46.
- Church, K., & Smyth, B. (2009). Understanding the intent behind mobile information needs. In *Proceedings of the 14th International Conference on Intelligent User Interfaces* (pp. 247–256). New York, NY: ACM.
- Cole, M.J., Gwizdka, J., Liu, C., & Belkin, N.J. (2012). Dynamic assessment of information acquisition effort during interactive search. *Proceedings of the American Society for Information Science and Technology*, 48(1), 1–10.
- Dankasa, J. (2017). Seeking information in circles: The application of chatman's life in the round theory to the information small world of catholic clergy in northern Nigeria. *Journal of Information Science*, 43(2), 246–259.
- Deerwester, S., Dumais, S.T., Furnas, G.W., Landauer, T.K., & Harshman, R. (1990). Indexing by latent semantic analysis. *Journal of the American Society for Information Science*, 41(6), 391–407.
- Eickhoff, C., Dungs, S., & Tran, V. (2015). An eye-tracking study of query reformulation. In *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 13–22). New York: ACM.
- Gwizdka, J., & Spence, I. (2007). What can searching behavior tell us about the difficulty of information tasks? a study of web navigation. *Proceedings of the American Society for Information Science and Technology*, 43(1), 1–22.
- Hansen, P. (2011). Task-based information seeking and retrieval in the patent domain : Processes and relationships. Tampere: Doctoral dissertation, Vaitoskirja Retrieved from <http://urn.fi/urn:isbn:978-951-44-8497-1>
- Ingwersen, P., & Jarvelin, K. (2005). The turn: Integration of information seeking and retrieval in context (the information retrieval series). Berlin, Heidelberg: Springer.
- Jansen, B.J., Spink, A., & Saracevic, T. (2000). Real life, real users, and real needs: A study and analysis of user queries on the web. *Information Processing & Management*, 36(2), 207–227.
- Jiang, J., He, D., & Allan, J. (2014). Searching, browsing, and clicking in a search session: Changes in user behavior by task and over time. In *Proceedings of the 37th International ACM SIGIR Conference on Research &#38; Development in Information Retrieval* (pp. 607–616). New York: ACM.
- Johannisson, J., & Sundin, O. (2007). Putting discourse to work: Information practices and the professional project of nurses. *The Library Quarterly: Information, Community, Policy*, 77(2), 199–218.
- Kellar, M., Watters, C., & Shepherd, M. (2007). A field study characterizing web-based information-seeking tasks. *Journal of the American Society for Information Science and Technology*, 58(7), 999–1018.
- Kelly, D. (2006a). Measuring online information seeking context, part 1: Background and method. *Journal of the American Society for Information Science and Technology*, 57(13), 1729–1739.
- Kelly, D. (2006b). Measuring online information seeking context, part 2: Findings and discussion. *Journal of the American Society for Information Science and Technology*, 57(14), 1862–1874.
- Kumpulainen, S. (2014). Trails across the heterogeneous information environment: Manual integration patterns of search systems in molecular medicine. *Journal of Documentation*, 70(5), 856–877.

- Kumpulainen, S., & Järvelin, K. (2010). Information interaction in molecular medicine: Integrated use of multiple channels. In *Proceedings of the Third Symposium on Information Interaction in Context* (pp. 95–104). New York: ACM.
- Li, Y., & Belkin, N.J. (2008). A faceted approach to conceptualizing tasks in information seeking. *Information Processing & Management*, 44(6), 1822–1837.
- Mitsui, M., Shah, C., & Belkin, N.J. (2016). Extracting information seeking intentions for web search sessions. In *Proceedings of the 39th international acm sigir conference on research and development in information retrieval* (pp. 841–844). New York: ACM.
- Orso, V., Ruotsalo, T., Leino, J., Gamberini, L., & Jacucci, G. (2017). Overlaying social information: The effects on users' search and information-selection behavior. *Information Processing & Management*, 53(6), 1269–1286.
- Pharo, N., & Järvelin, K. (2004, May). The SST method: A tool for analysing web information search processes. *Information Processing & Management*, 40(4), 633–654.
- Rose, D.E., & Levinson, D. (2004). Understanding user goals in web search. In *Proceedings of the 13th International Conference on World Wide Web* (pp. 13–19). New York: ACM.
- Saastamoinen, M., & Järvelin, K. (2016). Queries in authentic work tasks: The effects of task type and complexity. *Journal of Documentation*, 72(6), 1114–1133.
- Saastamoinen, M., & Järvelin, K. (2018). Relationships between work task types, complexity and dwell time of information resources. *Journal of Information Science*, 44(2), 265–284.
- Saastamoinen, M., Kumpulainen, S., & Järvelin, K. (2012). Task complexity and information searching in administrative tasks revisited. In *Proceedings of the 4th Information Interaction in Context Symposium* (pp. 204–213). New York: ACM.
- Savolainen, R. (1995). Everyday life information seeking: Approaching information seeking in the context of way of life. *Library & Information Science Research*, 17(3), 259–294.
- Silverstein, C., Marais, H., Henzinger, M., & Moricz, M. (1999). Analysis of a very large web search engine query log. *ACM SIGIR Forum*, 33(1), 6–12.
- Sohn, T., Li, K.A., Griswold, W.G., & Hollan, J.D. (2008). A diary study of mobile information needs. In *Proceedings of the sigchi conference on human factors in computing systems* (pp. 433–442). New York: ACM.
- Spink, A., Wolfram, D., Jansen, M.B.J., & Saracevic, T. (2001). Searching the web: The public and their queries. *Journal of the American Society for Information Science and Technology*, 52(3), 226–234.
- Teevan, J., Alvarado, C., Ackerman, M.S., & Karger, D.R. (2004). The perfect search engine is not enough: A study of orienteering behavior in directed search. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 415–422). New York: ACM.
- Teevan, J., Ramage, D., & Morris, M.R. (2011). #Twittersearch: A comparison of microblog search and web search. In *Proceedings of the fourth ACM International Conference on web Search and Data Mining* (pp. 35–44). New York: ACM.
- Vakkari, P. (2000). Cognition and changes of search terms and tactics during task performance: A longitudinal case study. In *Content-based multimedia information access - volume 1* (pp. 894–907). Paris, France: Le Centre de Hautes Etudes Internationales D'informatique Documentaire.
- Vakkari, P. (2003). Task-based information searching. *Annual Review of Information Science and Technology*, 37(1), 413–464.
- Veinot, T.C. (2007). "The eyes of the power company": Workplace information practices of a vault inspector. *The Library Quarterly*, 77(2), 157–179.
- Whitley, R., & Frost, P. (1973). Task type and information transfer in a government research laboratory. *Human Relations*, 26(4), 537–550.
- Xie, I., & Joo, S. (2012). Factors affecting the selection of search tactics: Tasks, knowledge, process, and systems. *Information Processing & Management*, 48(2), 254–270.